



2006

# Puppetry v. Creationism: Why AI Must Cross the Chasm

Hayes-Roth, F.

---

Hayes-Roth, F. (2006). "Puppetry v. Creationism: Why AI Must Cross the Chasm." *Intelligent Systems* 21(5) (Sept/Oct): 7-9.



Calhoun is a project of the Dudley Knox Library at NPS, furthering the precepts and goals of open government and government transparency. All information contained herein has been approved for release by the NPS Public Affairs Officer.

**Dudley Knox Library / Naval Postgraduate School**  
**411 Dyer Road / 1 University Circle**  
**Monterey, California USA 93943**

<http://www.nps.edu/library>



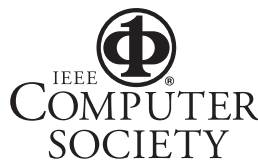
[www.computer.org/intelligent](http://www.computer.org/intelligent)

## **Puppetry vs. Creationism: Why AI Must Cross the Chasm**

**Rick Hayes-Roth, *Naval Postgraduate School***

Vol. 21, No. 5  
September/October 2006

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.



© 2006 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

For more information, please see [www.ieee.org/portal/pages/about/documentation/copyright/polilink.html](http://www.ieee.org/portal/pages/about/documentation/copyright/polilink.html).

# Puppetry vs. Creationism: Why AI Must Cross the Chasm

Rick Hayes-Roth, *Naval Postgraduate School*

**A**rtificial intelligence began with an enthusiastic embrace of newly available computing machinery and the basic question of what kinds of problems we could solve with it. The first 50 years focused on programming com-

puters to perform tasks that previously only humans could do. Then, people began comparing machines to humans as problem solvers, and the race was on to see where machines could match or even surpass human performance. Success in solving math word problems, winning checkers and chess championships, understanding natural language, and generating plans and schedules reinforced our efforts to build supercapable machines. I call these puppets, not to derogate the machines but to respect the importance of the programmers and builders who were actually responsible for their accomplishments.

From time to time, many of us have recognized the field's rate-limiting factor under various names and viewpoints, such as the knowledge-acquisition bottleneck and the challenges of machine learning, system bootstrapping, artificial life, and self-organizing systems. Some have focused on efforts to create a large corpus of off-the-shelf knowledge that would enable the next puppet to stand on the shoulders of its predecessors. Mostly, however, these efforts have had limited success. The little bit of learning and adaptation they've demonstrated has paled in comparison to the puppeteers' laborious inputs.

I believe we're on a local maximum, making better and better puppets with no apparent increase in speed or acceleration. We're stuck, and puppet making is a technology begging to be leapfrogged.

### Problem and analysis

The best systems of our times have been mostly hand-crafted by great engineers. These puppet makers have analyzed the task environments, knowledge requirements, and reasoning skills necessary for successful applications, and they've addressed them with better and better tools over time. This approach can work for any well-defined and sufficiently narrow task. If the puppets failed, the

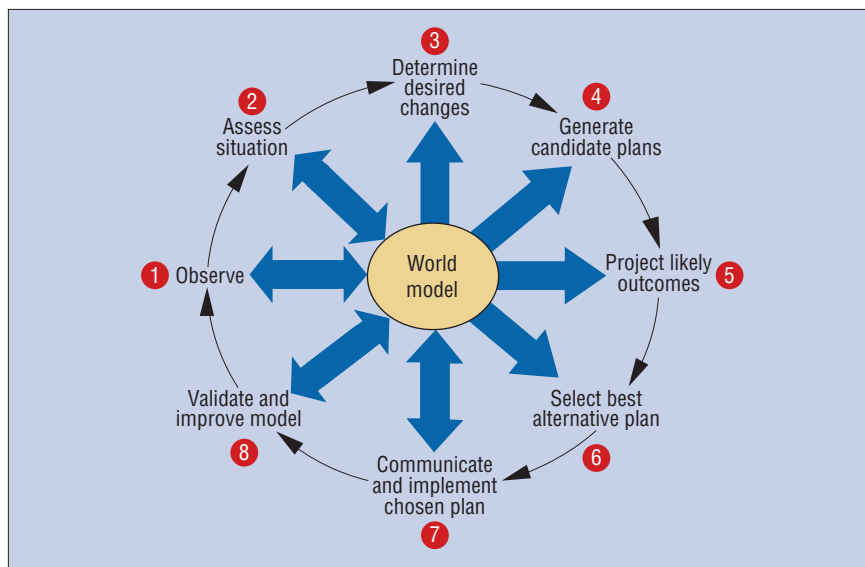
engineers would diagnose and debug the errors. They would determine what knowledge to add or modify, how to program it, and how to modify and rebalance the pre-existing programs to accommodate the new performance without harming the parts that already worked well. Automation in adaptation, learning, and knowledge acquisition was very limited—a tiny fraction of the overall knowledge required, which the engineers mostly prepared manually.

We haven't yet figured out how to make the puppets responsible for their own debugging and improvement. Because we're mostly labor intensive, we're on a curve of diminishing returns. Efforts to address this productivity decline through reusable knowledge bases have had limited success, chiefly because human engineers must comprehend the problem, analyze the knowledge requirements, determine how to adapt the available knowledge to the new application requirements, and conduct the essentially experimental cycle of modifying the implementation, then testing, diagnosing, refining the knowledge, and adapting the code. So, even though the puppets get more marvelous, the credit goes to the puppeteers. Moreover, the time intervals between significant improvements aren't decreasing. The rate-limiting factor is the speed with which human engineers can change the puppets.

### Vision and opportunity

In the dome of the Sistine Chapel, Michelangelo Buonarroti's fresco shows the creation of Adam with a pregnant touch between the hand of God and the hand of Adam. Many interpret this as illustrating the simultaneous animating of Adam and the bestowal of free will. In AI jargon, Adam was granted autonomy and agency, as well as responsibility. This is a bit of a conundrum for a naïve, ignorant, or merely inexperienced creature. Given the lack of knowledge and experience, such creatures are bound to commit errors. To thrive, they must incorporate a strong drive to improve as well as an effective process for continuous improvement.

Continuous improvement is widely taken for granted as the sine qua non of organizational excellence, but we can be pretty sure that the biblical creation story took a lot of this



**Figure 1. Efficient thought employs eight key functions supported by a world model.**

though no one has an accurate crystal ball, many visions are ultimately realized. Several long-range forecasts about AI have proved true. We have champion game players, autonomous vehicles, mobile robots, expert systems, speech transcription systems, and so forth. In other cases, practice has come up short, or the visions themselves have turned out to be absurd.

Most of the visions for self-modifying and self-improving AI have been a bit scary or a bit shallow. Much discussion has occurred around the concept of the *singularity*, first explained by Vernor Vinge<sup>2</sup> and now a theme popularized by Ray Kurzweil,<sup>3</sup> among others (for example, see [www.aleph.se/Trans/Global/Singularity](http://www.aleph.se/Trans/Global/Singularity)). Roughly, the singularity is a point in history when technology accelerates beyond human capacity to master it. With regard to AI, this could mean that computers learn and communicate with one another faster than they can with humans. At that point, they might not be willing to slow down or engage further. In such a case, the machines would reach a kind of escape velocity, enabling them to leave human culture behind.

I'm suggesting that we should actively seek to create the capabilities underlying such a possibility, because the potential gains are exponentially greater than what traditional puppetry can produce. But this surely entails risks, as readily suggested in most of the scary movies about rogue AI and robots gone haywire.

We obviously have limited insight into both the positive and negative capabilities of self-improving AI systems. Although the road will be somewhat long, we ought to consider ways to identify and mitigate the risks before they afflict us. Fortunately, many people interested in the singularity and future visions of robots have given these issues serious consideration. We'll want to incorporate their ideas and related technologies into the new creationism agenda. This will help prevent predictable problems and provide additional insurance against the unforeseen.

Although I consider these risks serious, I think we can de-emphasize or defer them a bit. To get out of the puppetry business will require a major shift in investment, orientation, and technology. By focusing mostly on the required new capacities for model-based learning and improvements in operational contexts, we put first things first.

for granted. Nevertheless, the idea that creation would launch many sentient, autonomous, continuously improving, and responsible creatures suggests the way forward for AI. If we want our puppets to fulfill AI's potential, we need to launch a new category of critters that have similar capabilities. Making better and better clockwork puppets won't do it for us; we simply can't evolve them fast enough.

Many researchers have gone after this idea by focusing on simple creatures that could adapt and evolve through relatively simple, mechanistic techniques. They're making good progress, albeit low on the evolutionary tree. We should be launching our evolutionary efforts on a much higher plane of understanding and sophistication so that we can avoid the long process of recreating the kind of human knowledge that engineers now regularly transfer into puppets.

I believe the high-reward, low-risk approach is straightforward. We should shift most of our R&D to self-improving, high-competence, knowledge-intensive systems. We should launch a new age of creationism, borrowing from the creation of Adam to focus on sentient, intelligent, self-improving, and responsible agents.

## Efficient thought as a blueprint

We want critters that can plan and act in the world with continuously improving results. These agents will make many decisions based on their beliefs about how the

world works, which I term their "world model." That model lets them interpret observations by instantiating their parameterized models to match the observations. In other words, the agents can perform analysis through synthesis. Their world model lets them predict likely outcomes of actions and dynamic processes by computing the implications of hypothetical model states. With this capability, the agents can choose promising plans by selecting those that lead to favorable predicted outcomes.

I call this entire cycle *efficient thought*.<sup>1</sup> Figure 1 illustrates it in eight steps numbered in a typical sequence, although most complex organizations perform all eight steps in parallel. The intelligent being (1) observes what's happening in the environment, (2) assesses the situation for significant threats and opportunities, (3) determines what changes would be desirable, (4) generates candidate plans for making those changes, (5) projects the likely outcomes of those plans, (6) selects the best plan, and (7) communicates that plan to key parties before implementing it. Throughout the process, the intelligent being (8) validates and improves its model. The model supports all eight activities, although only steps 1, 2, 7, and 8 directly update and modify the model.

## The singularity and its risks

Futurologists, science fiction writers, and other visionaries often foresee events. Al-

**L**ike the master puppet makers of the classic folktales, AI engineers have built some marvelous machines. These machines are surely valuable, but our puppetry won't cross the chasm separating us from a world of artificially intelligent creatures. To get there, we need a singularity of artificial creationism, where we launch artificial beings that can adapt, learn, and evolve. We need to emphasize the development of continuously self-improving systems that interact with and perform tasks in the physical world. Creation of those systems will mark a singularity in the punctuated evolution of artificial intelligence. ■

## References

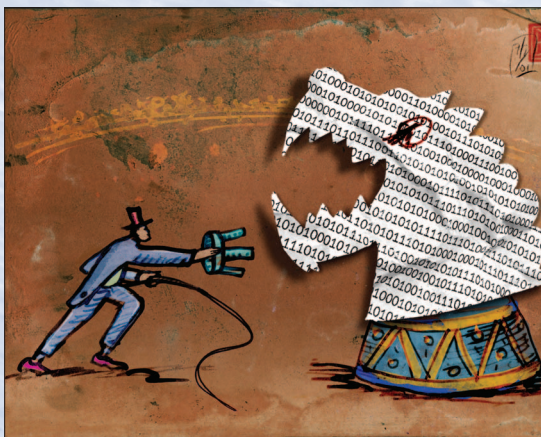
1. R. Hayes-Roth, *Hyper-Beings: How Intelligent Organizations Attain Supremacy through Information Superiority*, Booklocker.com, 2006; [www.hyper-beings.com](http://www.hyper-beings.com).
2. V. Vinge, "Essay on the Singularity," presented at VISION-21 Symp., 1993; <http://mindstalk.net/vinge/vinge-sing.html>.
3. R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, Viking, 2005.

For more information on this topic or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).



**Rick Hayes-Roth** is a professor in the Information Sciences Dept. at the Naval Postgraduate School in Monterey, California. Contact him at [hayes-roth@nps.edu](mailto:hayes-roth@nps.edu).

## Building the community of leading software practitioners



# Editorial Calendar 2006–2007

## IEEE Software

[www.computer.org/software](http://www.computer.org/software)

**NOVEMBER/DECEMBER**  
Software Engineering Curriculum Development

**JANUARY/FEBRUARY**  
Software Engineering Challenges  
in Small Software Organizations

**MARCH/APRIL**  
Needs-Driven Dependability Engineering

**MAY/JUNE**  
Requirements Engineering: Understanding  
the Stakeholders' Desires and Needs

**JULY/AUGUST**  
Test-Driven Development

**SEPTEMBER/OCTOBER**  
Rapid Application Development  
with Dynamically Typed Languages